

Abstract

A system and methods are provided for automatically classifying data according to perceptual properties of the data to form a classification chain that is suited to the searching and sorting of large databases of media entities. During classification, experts assign each media entity in the training data set to one or more classes, with each class corresponding to a given subset of perceptual properties of the data. In conjunction with digital signal processing properties of the data corresponding to the perceptual properties, the classified data is then used to construct an initial classification chain. During operation, when presented with an unclassified entry, the classification chain returns an estimate of the class of the entry, as well as a confidence measure that is proportional to the level of confidence of the class assignment. Over time, as the classification chain evolves, the classification chain becomes more and more effective for quickly characterizing media entities.